DOCUMENT RESUME

ED 318 136 EC 230 087

AUTHOR Callahan, Carolyn M.
TITLE Program Evaluation as an Essential Component of Programs for the Gifted and Talented. Monograph Prepared for the Leadership Accessing Program.
INSTITUTION Indiana State Dept. of Education, Indianapolis. Office of Gifted and Talented Education.; Purdue Univ., West Lafayette, IN. Gifted Education Resource Inst.
PUB DATE 88
NOTE 25p.
PUB TYPE Guides - Non-Classroom Use (055) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Gifted; *Program Evaluation; *Research Problems; *Talent

ABSTRACT

A rationale is presented for the systematic evaluation of gifted/talented education programs, citing the consequences of failure to carry out such evaluation. Historically, evaluation of gifted programs was not emphasized, as the evaluation process involved experimental designs, control groups, and parametric statistics which could only be understood by select statisticians or experimental psychologists. A new subscience has emerged which focuses not only on student outcomes but also on program delivery processes, program milieu, resource inputs, and other factors. Fourteen pitfalls to be avoided in the program evaluation process are explored. These pitfalls include failing to recognize assumptions, avoiding "hard" or "uncomfortable" issues, and assuming that qualitative evaluation will be easier than quantitative evaluation. (14 references) (JDD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EC

ED318136

PROGRAM EVALUATION AS AN ESSENTIAL COMPONENT OF PROGRAMS FOR THE GIFTED AND TALENTED

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BY

CAROLYN M. CALLAHAN, PH.D.

UNIVERSITY OF VIRGINIA

A MONOGRAPH PREPARED FOR THE LEADERSHIP ACCESSING PROGRAM

INDIANA DEPARTMENT OF EDUCATION
OFFICE OF GIFTED AND TALENTED PROGRAMS
PATRICIA B. STAFFORD, MANAGER

IN CONJUNCTION WITH

PURDUE UNIVERSITY
GIFTED EDUCATION RESOURCE INSTITUTE
JOHN F. FELDHUSEN, DIRECTOR

JULY 1987

EC 230087

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Patricia B. Stafford

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Copyright 1988

Indiana Department of Education
H. Dean Evans, Superintendent

INTRODUCTION

In this monograph on "Program Evaluation As An Essential Component Of Programs For The Gifted And Talented," Carolyn Callahan tells us the many mistakes we are likely to make concerning evaluation of gifted programs, and she points the way to proper, well-conceived evaluation. While gifted programs have grown by leaps and bounds, our efforts to assess the value of those programs have developed slowly. Sometimes program leaders are defensive and don't really want to know (or fear) the truth; other times they are simply so weak in their knowledge of methods for program evaluation that they take no action.

Traditionally, Callahan points out, the process of evaluation was more or less synonymous with experimental design. Thus, there were experimental designs, control groups, parametric statistics, and a general evaluation milieu which could only be understood by a select few statisticians or experimental psychologists. Gradually, however, a new subscience emerged with a unique set of methodologies for program evaluation. These new designs focused not only on student outcome variables as the bottom line of evaluation methodology, but also on program delivery processes, the context or milieu in which the program was delivered, and on the input of resources incorporated into the program. The latter would include the goals and objectives of the program, as well as the specific nature or characteristics of the gifted population served.

Thus, the overall context of gifted program evaluation has been enlarged considerably. Descriptive and inferential research design models are used. Qualitative designs are also considered because they can provide a richer pattern of outcomes or effects. Evaluation of gifted programs and of curriculum for the gifted has come of age. The problem now, according to Callahan, is to convince all interested parties of the relevance and value of sound program evaluation.

BRIEF BIOGRAPHY OF THE AUTHOR

Carolyn Callahan received her Doctorate from the University of Connecticut in 1973. Since that time, she has been employed by the University of Virginia where she is currently a Professor in the School of Education and Director of the Summer Enrichment Program. Her publications include: *Developing Creativity in the Gifted and Talented*, "Issues in Evaluating Programs for the Gifted," "Gifted Women," *The SEA Test*, and *New Directions in Creativity: Mark III*. A past-president of the Association for the Gifted (TAG), Dr. Callahan currently serves on the Board of the National Association for Gifted Children, and also serves on the editorial boards of *Roeper Review, Exceptional Children, Gifted Child Quarterly,* and is book review editor for the *Journal for the Education of the Gifted.*

PROGRAM EVALUATION AS AN ESSENTIAL COMPONENT OF PROGRAMS FOR THE GIFTED AND TALENTED

INTRODUCTION TO THE EVALUATION ISSUE

How do you know that your gifted program is not causing self-concept problems for those students who are not identified as gifted?

Does the program just take the good students out of the classroom and remove the stimulation normally given to other students?

How do you know that the students are learning things that they would not learn if the program did not exist? How do you know that the things they are learning are worth learning; are the most important things for them to learn, and are taught efficiently and effectively?

How do you know that this program is worth the cost in student time, teacher time, aggravation, scheduling problems, etc?

How do you know that the students are not missing some important skill development that is part of the regular curriculum?

Everyone who has been involved in developing programs for the gifted and talented has faced each of these questions (and many more) about gifted students and gifted programs at some time. They are questions which are very frequently raised as gifted programs are developed and implemented. They are usually answered with general statements about the importance of not boring gifted children, with warnings about the danger of gifted dropouts, with statements about the negative consequences of not having a program, etc. Unfortunately, program developers quickly forget those questions once the program has been established and the day-to-day operational issues must be confronted -- forget them until some unhappy parent, disgruntled administrator or teacher, or fiscally concerned board member raises the same questions again. Then lack of attention to evaluating the gifted program and its goals leaves many educators with no response to such questions, and gifted programs become vulnerable to criticisms. *Even more unfortunately*, neglect of evaluation often means that the programs that have been offered to the gifted have been less adequate than they might have been. In some cases, they have even been completely inadequate for some of the gifted students we seek to serve. How can advocates of appropriate education for gifted (or all) students be

satisfied with providing a program which is less than optimal? And how can administrators be satisfied until they can be assured that the expenditure of time, money, and effort is delivering the best program possible?

Look at this issue from another point of view. Examine the list of goal statements given below:

*The development of logical reasoning abilities
*The development of critical thinking skills

    *judging credibility of sources
    *making judgments based on unbiased evidence

*The development of analytical reasoning
*The development of analysis and evaluation skills

    *making judgments on the basis of internal evidence -- evaluation on the basis of accuracy logic, and consistency
    *making judgments on the basis of external criteria -- evaluation of materials with reference to standards and criteria

*The development of creativity

Do these goals look familiar? Each of these statements was taken from a list of goals and objectives of a gifted program and is probably similar to statements of goals or objectives found in ninety percent of the gifted programs in the United States. It is paradoxical that programs for the gifted can focus so intently on the development of logical thinking, stress the need for the collection of systematic evidence upon which to make judgments, urge students to be critical judges of the environment in which they live, promote creative solutions to problems, *and yet fail* to apply these same principles to the very programs in which the gifted students are placed. It seems that educators are willing to accept on face value or at "gut level" the assumptions that gifted programs are serving the right students, that the programs are meeting the needs of the identified students, that their programs are the most appropriate programs for the identified students, that stated goals and objectives are being achieved as a result of the gifted program, and so on.

Consider a medical analogy. In modern medicine, before a treatment or a drug is widely disseminated and/or prescribed, physicians and patients seek proof that the treatment is effective. In contrast, in the case of gifted and talented education, once the determination has been made that a program is needed, the assumption is generally made that a given program will be effective, even in the face of the fact that precious little, if any, evidence exists to support such an assumption. In medicine, consumers ask whether a given treatment or drug will be the most effective, most cost efficient, and least dangerous. Patients are concerned about possible side effects, about indicators that the drug or treatment is actually treating the symptom and/or the disease, and about insuring that the individual who is administering the treatment has the skills to deliver, monitor, and assess the effects of the intervention appropriately. Sadly, by

contrast, in the field of gifted education, little attention has been given to systematically collecting data about the "treatments" offered to our students -- about the program's effectiveness, its side effects, the personnel who deliver the instruction and their qualifications, etc. Finally, in the medical model, a physician looks carefully at the effects of the treatment on the *individual* to whom the treatment is offered regardless of what group data indicates. How often in gifted education do program personnel justify the implementation of a given model of gifted education on the basis that it worked somewhere else, or "the literature says it is best" without examining how effective it is in a *particular* school system, for *particular* students, etc.? Unfortunately, the answer to this question is: *Far too often*. In fact, most gifted programs fail to evaluate at all, and when they do, they focus primarily on ill-constructed surveys of attitudes and perceptions (Gallagher, Weiss et al, 1983).

The dearth of evaluations which go beyond the collection of data relating to student, parent, teacher, and administrator perceptions of program success is increasingly embarrassing as the growth of gifted programs continues at a rapid pace. The proliferation of programming models continues with little attempt to assess systematically the logical consistency of programs within their settings to evaluate the degree to which programs meet the particular needs of a given school, to discover the degree to which particular models are being implemented as designed, to measure the "fit" of given models in given schools situations (school environments), to judge the effectiveness of these models in achieving the particular goals and objectives of individual school corporations, to weigh the impact of the gifted program on other school programs, etc. According to Gallagher, Weiss et al (1983),

"The data collected on program evaluations for gifted programs suffer from a lack of attention to evaluation design and a lack of trained evaluation personnel. More and better studies need doing before confidence can be placed in these results. The authors believe that there should be much more work done on the...development of task analysis of skills such as independent study in order to better assess student entry and exit levels, to demonstrate useful products as possible outcome measures, to demonstrate the generalization of learned skills, and to have more attention paid to basic research and evaluation design. The area of gifted education remains a fertile and largely unexplored field in need of consistent formative evaluation." (pp. 69-70)

In light of this lack of evaluation data and a lack of "basic research" in the area of gifted education (and the lack of attention to the evidence that does exist), one cannot help but wonder at the confidence with which school personnel ask for monies to expand programs. And, one cannot help but wonder when someone will begin to hold educators accountable for how those funds have been expended. That time is bound to come, and without serious attention to evaluation the field of gifted education and particular programs are likely to be sadly lacking in evidence of effectiveness. But more importantly, aren't *gifted educators* interested in knowing exactly what impact gifted programs are or are not having on gifted students? Aren't educators interested in identifying which components of gifted programs are functioning appropriately and which are not? Aren't educators concerned about the possibility that gifted students have developed an elitist attitude? Isn't the goal of any educational

program the provision of the best possible program? How can that best possible program be delivered without assessment of the degree to which our goals are met, the degree to which the program is actually being delivered, etc.?

HISTORICAL PERSPECTIVE

The lack of emphasis on the evaluation of gifted programs is easily understood if one considers the history of evaluation, the history of gifted programs, and the way in which the two fields have related to one another (or failed to relate to one another). First, if we examine the field of evaluation, we find an evolution of the evaluation process from focus on the traditional experimental design paradigms (through the mid 1960s and early 1970s) to a focus which currently incorporates experimental design as a component only when certain preconditions and evaluation questions warrant that focus.

If we examine the development of gifted programs during the same time frame, we see that many early programs were developed in the 1960s and 1970s with the use of federal funds through monies allocated for experimental programs directly from the U.S. Office of Gifted and Talented or indirectly through Title IV-C grants. These grants always were accompanied by mandates for evaluation as a condition for funding and accordingly were scrutinized by evaluators or researchers who were still almost exclusively concerned with outcome data judging the worth of programs (and consequent implications for continued funding). These individuals were seldom concerned about evaluative data which could yield information on the processes and products of the program and the factors which might be contributing to or detracting from program effectiveness. In addition, these evaluators were most often engaged in evaluating programs focusing on nonachieving students. Consequently, evaluation was likely to consist of attempting to establish experimental or treatment groups and control groups, picking an objective instrument presumed to measure important outcomes of a gifted program, and then setting up pre- and posttesting times. Little or no attention was paid to some of the most obvious problems with such an approach.

One obvious problem was that there was little resemblance between the goals of gifted programs and those of a program for average or slow learners -- and there should not be. Therefore, the existing objective instruments were inappropriate at best and ridiculous at worst. Incorrect assumptions about the concurrence between curricular efforts for the gifted, the outcomes measured by standardized achievement tests, and unverified assumptions about the needs of gifted students resulted in the use of existing standardized achievement tests to measure the outcomes of gifted programs. Lack of understanding about the characteristics of the gifted led evaluators to assume that a regular classroom could serve as control group. Lack of understanding of terminology led to the choice of inappropriate instruments -- for example, the use of reading comprehension tests to assess analysis skills or the Otis-Lennon Intelligence Test to measure higher level thinking skills.

In addition to the problems mentioned above, the evaluation often focused on the need to convince others of the worth of particular programming strategies, curricular models, or instructional practices so that these models could be adopted or replicated in other settings (or at least get reported in a journal). Thus the interest was *not* on using evaluation results to improve the program, *but* rather as a means of "certifying" or eliminating the program. And, the emphasis was not on the issues of program improvement or on the appropriateness of the

context of the program as much as on the determination of the generalizability of findings. Evaluation in the field of gifted education often came to be perceived as an expensive process which only made summative value judgments on the outcomes of instruction.

With the demise of the U.S. Office of Gifted and Talented and Title IV-C, gifted programs were funded largely through school district allotments or state reimbursement plans. Receipt of these monies seldom required a formal evaluation of programs, and when evaluations were mandated, a survey of parent, student and/or teacher perceptions usually sufficed as evidence of success. School corporations themselves were not motivated to look beyond these data for several reasons. First, the implementation of an evaluation plan is often perceived as very costly. Even if only a small segment of a program is evaluated and that evaluation is done internally, there *are* costs in terms of personnel time, student time and money for tests or test development, surveys, data analysis, mailings, etc. That cost was not viewed -- because of earlier experience -- as a means of program improvement. Second, there are often no individuals on a school staff who have had experience conducting evaluations, let alone an evaluation of a gifted program. Only very large schools employ a research and evaluation staff and their responsibilities are often directed at the monitoring of regular school programs and large internal or external funding efforts. Small schools do not even have staff with training in the area of evaluation. In addition, individuals who are selected to direct gifted programs in smaller schools are often not only administrators of gifted programs, but are in charge of several other programs as well and have no time even to consider evaluation.

The alternative for these programs is external evaluation. Unfortunately, if the evaluation is done by an external evaluator, there are additional evaluator/consultant costs. Because the budgets of gifted programs are often so limited, finding the additional funds to support evaluation may be very difficult. Add to all of these considerations the suspicion that most people have that an evaluation will probably do the program no good and could potentially point out problems that someone would rather keep under wraps, and it is not surprising that very few individuals at the school district level ask for resources to evaluate their gifted programs. Nonetheless, the recent attention to "accountability" reports that our education programs are inadequate (consider recent news reports that our math instruction is the *worst* of all industrialized nations), and the "coming of age of gifted programs" have led to a time when questions about effectiveness, efficiency *and even* appropriateness are inevitable. The time has really come when educators can no longer afford *not* to evaluate gifted programs. For, if the field of gifted education allows itself to become satisfied with current efforts, based on little or no evidence of the impact of the programs offered, the functioning of the programs, or the degree to which the program is being implemented as described, then schools have failed in their mission to provide the best educational program for gifted students, or even students in general. Because it is certain that no gifted program is yet "perfect" or even "as good as it can be in this school system," it is also certain that program administrators need data to determine what aspects of programs are functioning appropriately and what aspects of programs need improvement, refinement, expansion, elimination, and revision. Yet, as administrators accept responsibility for examining gifted programs, it is quite possible that they are ineffective in producing useful evaluations unless they watch for "Pitfalls of the Unwary Evaluator."

WHAT ARE THE PITFALLS THE UNWARY EVALUATOR SHOULD AVOID?

Even in those cases where school administrators or evaluators have accepted the need to conduct an evaluation of the gifted program, there are many reasons why the "science of evaluation" has not been practiced as effectively in evaluating programs for the gifted as one might hope. These reasons might best be expressed as a failure to answer completely the questions: "Evaluation for what purpose?" and "Evaluation for whom?" Moreover, in many instances, these questions have been neglected altogether. As a result, program administrators and teachers of the gifted continue to do something because they always have, or something new just because it feels good or seems right and everyone else is doing it, regardless of the true purpose of the activity.

To address questions of evaluation purpose and appropriate audiences adequately and still not be able to answer them completely (these questions are, after all, not always as easy to answer as they appear at first blush) will somewhat restrict the utility of the resulting evaluation. Such restrictions are unfortunate, but *not nearly* as *unfortunate* as the current situation in which the questions of purpose and audience are not considered at all or given only cursory attention Thus, it seems most appropriate to offer guidelines to be used in addressing these questions, knowing that the degree to which a school system addresses them will greatly influence the degree to which the evaluation will serve the function of program involvement and justification.

Pitfall 1: Failing to Consider the Range of Answers to the Question, "Evaluation for What Purpose?"

The reasons for or the purposes of evaluation fall into three basic categories. As mentioned above, traditional evaluations of gifted programs have historically been for the purpose of convincing "potential adopters that a particular approach or model works and can be generalized to other settings (Category 1). That approach has been emphasized in programs which are designed to be models for other programs or projects. Evaluation might also be done to convince some audience (school board, administrators, funding agents, parents, etc.) that a program "works," and that, therefore, it should be continued (Category 2). Finally, evaluation can be used to determine how well a program is working, what its strengths and weaknesses are, and how it can be made to work better (Category 3).

For two and a half decades, program evaluators have been arguing the relative merits of evaluations performed primarily for the purpose of providing information for program improvement (Category 3 above) against the merits of evaluations performed primarily for the purpose of providing information for making judgments about the "worth/merit" of a program (Categories 1 and 2 above). While both of these purposes are certainly legitimate, evaluation for the purpose of program improvement is often considered to be "useful and credible" only to program staff and is, therefore, relegated to the category of a program

development activity. Evaluation information focusing on worth or merit, on the other hand, is desired and most often demanded by those individuals or groups who make decisions about the fate of gifted and talented programs; i.e., to continue to fund or to terminate. It is most beneficial to programs to accept both of these purposes of evaluation as legitimate. The question is not one of choosing one purpose or another purpose, but rather one of asking which purpose of evaluation will best serve the program, the program consumers (i.e., students), or the decision-makers at a given point in time. Consider the range of possible purposes presented by Callahan and Caldwell (1986):

Documentation of the need for a program.

Documentation of the case for a particular approach.

Documentation of the feasibility of a program.

Documentation of program implementation.

Identification of program strengths and weaknesses.

Provision of data for in-progress revisions of the program.

Documentation of the results or impact of a program.

Explanation and description of the program to interested but uninformed audiences.

Pitfall 2: Failing to Select an Appropriate Purpose in Evaluating a Gifted Program

In selecting the purpose or purposes of an evaluation of a program for gifted and talented students, there are several factors to be considered. One of the factors which must be considered is the "age" or stage of development of a program. The field of education of the gifted is a relatively new field and many programs are truly experimental in nature. The research base is scanty at best; there are many opposing definitions of giftedness, divergent theories of curriculum, and conflicting outcome statements that compete for primacy in program development. Thus, at an early stage of program development in a given setting, it is premature to consider the evaluation of student outcomes. In the early stages of program development, the evaluation should focus on logical consistencies among needs of the children to be served, program prototype selected, curriculum to be offered, and staff training. Similarly, as a program begins to mature, the major focus of evaluation shifts to an evaluation of consistency between stated program functions and school implementation, to a measure of success of training, to assessment of the adequacy of curriculum development, and to judging the degree to which curriculum is actually implemented. Finally, as a program has matured, one may begin to examine student outcome data. (Interestingly, student objectives often change as programs mature and as administrators become more sophisticated about the field.) This is not to suggest that an administrator should necessarily ignore student outcomes in early

program development and evaluation. Rather, it is important that the program evaluation not rest solely on those data. For, of what use is it to know that students have not achieved the stated goals of a program if one cannot determine the reason why?

Another set of factors associated with the selection of the purpose of evaluation relates to the nature of these gifted programs. First, one should consider the fact that the gifted and talented students will continue to exist, to have educational needs, and to be part of the school population, whether a given program exists or not. Second, there is no universal definition of giftedness; each group of students within a given school district is likely to be unique, as is the setting in which the gifted program operates. Third, a gifted program always operates in a highly political milieu. That is, the program is probably the result of pressure from outside agents such as parents or state law/mandate/standard and probably faced some degree of resistance in its inception. It is probably still regarded as a frill, an elitist activity, and/or a drain on regular programs (good students drained off to other classes, good teachers drained off to other programs, and good money drained off to an unpopular cause). Unlike a new reading program which addresses objectives with universal appeal, a gifted program must defend its purposes, its activities, *and* its outcomes. Do teachers and parents *really* want critical thinkers who will question, change, and debate parents' values?

In addition, as the questions of evaluation are framed, the evaluator must be aware that the data collected may be used for political, as well as evaluative arguments. These factors suggest that evaluations of gifted programs need to address questions of improving programs, identifying strengths and weaknesses, and identifying factors which contribute to successes and failures of the program, rather than questions relating to continuation of abolition of entire programs. Evaluations will be most useful to the population served (gifted and talented students) if they frame questions in terms other than, "Should there be continued funding for gifted programs?" or "Should the gifted program be continued?" to quote Michael Patton (1981), "When a program is unjustly done in by an evaluation, or when a policy is unjustly changed, the victims care not whether it was by malice or incompetence." (p. 72)

In implementing an evaluation of gifted programs, there is a danger of losing sight of exactly which of the purposes mentioned above was selected as the most appropriate purpose for evaluation. Even if a clearly specified purpose was stated at the onset ("The school board wants to know whether our existing program is making a difference in the achievement of gifted and talented students"), it is very easy for those involved in the evaluation design to lose sight of the original purpose and change the parameters of the evaluation during the planning and implementation stages. For example, it may not be comfortable, easy, or, in some cases, even possible for the program staff members to state specifically the desired student outcomes of the program, so they turn to process evaluation rather than product evaluation. "Are the teachers using appropriate strategies for instruction?" rather than "Are students developing research skills?" In some cases, the original purposes may be lost because of a fear that if the original purpose is addressed, the program will not fare well under scrutiny of outcome measures. The responsibility of the evaluator is to ensure that difficulty of assessment, fear, or even duplicity do not block attention to the stated purposes of evaluation.

One potentially undesirable consequence of confusion over the intended purpose is the selection of inappropriate evaluation approaches, or worse, to design and implement evaluations which attempt to be all things to all people. The latter usually occurs when evaluators say, "Since we're doing this anyway, why don't we ask..." and, consequently, try to design the evaluation so that it attempts to answer any and all questions that might be raised by any audience. Given that evaluations are costly, the net effect of such an approach is most often to provide "weak" information to a host of questions. Obtaining "strong" answers to a few, well-stated questions of major importance is the more appropriate approach.

Pitfall 3: Failing to Address the Appropriate Audience

Associated with the confusion relative to the purpose(s) of the evaluation is the selection of the wrong audiences to consult in the selection of evaluation concerns and the reporting of evaluation results. For example, if the school superintendent mandates an evaluation of a program, then the concerns of the school superintendent should be paramount in the planning of the evaluation. If the State Department of Education requires evaluation data, then the State Department should be consulted about the kinds of evidence that will be acceptable to answer concerns of that Department. Too often, one audience calls for the evaluation and a second group or individual defines the questions of the evaluation. A superintendent may be interested in evidence that teacher in-service is effective in changing teacher behavior, while the program director interprets this in such a way that the evaluator collects data showing that participants judged the speakers to be well-prepared. It is very important in the process of evaluation to "go to the horse's mouth" to ensure that the decision-making audience's concerns have been accurately reflected in the planning of the evaluation.

Similarly, confusion over the purpose of evaluation and the consequent confusion over which audience is the appropriate audience will lead to inefficient, confusing, or inadequate reporting strategies. Ideally, an evaluation design will focus on answering questions which are of concern to "consumer(s)" of the evaluation, and the identification of these consumers should be paramount in the early design of the evaluation and the reporting of results. Use of appropriate data analysis techniques will ensure that parents, teachers, school boards, and superintendents of schools can understand and be able to interpret a document or even an oral presentation. Terms like "multiple-R," "cross-lagged panel design," "analysis of covariance," may not be familiar to them. In such instances, they come to regard evaluation as a mystical, magical process which somehow loses the sense of the humanity of the programs evaluated. As a consequence, they may either put too much faith in "statistics" or disregard what they can not understand. It is important that we not overcomplicate *or* oversimplify our evaluation questions or designs. We must clearly identify the process, the results, and the implications in language which can be understood by those who must use the results in decision-making.

Pitfall 4: Providing Too Much Information Too Late

Once the appropriate audience has been identified, the format and timing of evaluation reporting should be considered. The most common and least effective approach to evaluation reporting is a written document (often quite thick) presented at the end of a project period;

i.e., end of the year or end of the funding period. Some would argue that this is not a sound reporting approach because it places the decision maker (who is facing deadlines) in the position of having to digest a great deal of material, much of it unfamiliar, before making a judgment within a very limited time frame. In many cases, the report is too late to have an influence on the decision-making process at all. When determining audiences, one must also ask *when* data will be useful to the consumer of the report. Brief, periodic, oral, informal reports to decision makers are much more likely to achieve the desired effects. However, if the traditional end-of-project reporting approach is to be employed, the evaluator should provide interim reports and/or executive summaries which highlight the major foci of the evaluation.

Pitfall 5: Setting Inappropriate Levels of Rigor

Failure to define the purpose of evaluation may also lead to confusion regarding the necessary "rigor" to be incorporated in the evaluation design. In general, the more rigor that is required in a design, the more costly the design -- both in terms of resources used and in terms of potential dysfunction relative to the program being served. For example, a rigorous design in the traditional experimental research sense would require the random assignment of gifted students to control and experimental groups.

Even though there is the risk of being labeled "unscholarly" by colleagues, evaluators should be aware that the requirement of rigor in program evaluations is relative. The degree of required rigor is closely associated with the risk involved in making an error in the evaluation. While the preference is a design as scientifically rigorous as the context allows, it is more important that the design be (1) most rigorous when the risk involves promoting an ineffective approach or model throughout the gifted community, (2) less so, when the risk involves assessing the impact of a reduction in staff on an existing gifted program, and (3) even less so, when the risk involves a short term, small internal management plan change for a single student in a single program. Rigor is not to be abandoned in any of the aforementioned cases; however, the cost of rigor must be balanced against the degree to which persons using the data wish to generalize findings.

Pitfall 6: Failure to Define Giftedness Carefully

While the preceding discussion has focused on problems which are considered general to the field of program evaluation, there are many evaluation problems which are unique to programs for the gifted and talented. Despite recognition of these problems and issues for many years (Renzulli, 1975; Callahan, 1983), they remain unresolved. The first of these issues faced in the evaluation of gifted is a lack of a common definition of giftedness and, perhaps more serious, identification procedures which often belie the definition of giftedness adopted by the program. An evaluator must first determine the nature of the population which is actually being served and never assume that the group meets textbook definitions of giftedness. An evaluator may find that a plan to assess the effectiveness of a program for gifted students is meaningless because the population identified is not the one presumed to be served.

Pitfall 7: Making Unwarranted Assumptions or Failing to Recognize Assumptions

Any discussion of the process of evaluating programs for the gifted should begin by examining underlying assumptions and verifying assumptions. About the only safe assumptions one can make include:

-- Gifted students do exist in this school system and we can find them. (Do not assume everyone in every school believes this.)

-- The "regular" or traditional "age/grade level" curriculum in a school does not meet the total scope of needs of gifted students. The regular curriculum is in some way not meeting the total educational need of the exceptionally able student.

-- Someone has developed an alternative educational program for the gifted (modifying content of the curriculum, teaching/learning processes, and/or environment) which is presumed by someone in the school to meet the particular educational needs of these students.

What we do *not* assume in program evaluation is as important as what we do assume. For example:

-- Do not assume that the gifted have, in fact, been identified appropriately.

-- Do not assume that the particular program which has been implemented is the most appropriate for those identified.

-- Do not assume that all aspects of the regular classroom instruction were/are inappropriate.

-- Do not even assume the given program was needed by the students who have been identified as gifted.

These assumptions may well be evaluation *questions* that are overlooked and yet explain the lack of success of a gifted program.

Pitfall 8: Lack of Clear and Uniform Statements of Program Goals

Not only is the field of gifted education characterized by the use of broad terms such as "thinking skills," "research skills," etc., but the lack of clearly identified and specified curricular goals may result in a situation in which individual teachers in a program hold widely different notions of the goals of the program -- dependent on the interpretation given at the most recent G/T conference attended. These interpretations are not unwarranted, given the lack of precision associated with the statements of program goals and objectives, but they can result in gross errors in the selection of outcomes to be assessed, in determining appropriate means of assessing those outcomes, and in constructing measures of outcome goals.

Pitfall 9: Failing to Select Appropriate Measures

Even when program goals *are* clearly and accurately stated the kinds of goals and objectives written for gifted students, for the training of teachers, and even for program implementation, are not easily assessed with standardized tests. First, student goals are most often in the areas of higher level thinking skills, creativity, understanding the methodology of the discipline, and producing real-life products. Few instruments exist to assess these types of goals and even those which do exist are of limited utility, given the narrowness of their range of assessment. Some goals of gifted programs focus on achievements in the talent areas of art, music, drama, and leadership -- areas in which assessment of goals is extremely difficult.

Many of the goals of gifted programs are long term in nature (e.g., the development of self-directed learning skills), and difficult to segment into individual components. It is difficult to assess components and then put the components back together again for a composite assessment. One example which is widely cited is the construct of creativity. There are measures of fluency, flexibility, originality, elaboration, certain process dimensions, and personality factors, but when assessments of them are added together, it is doubtful that "Creativity" with a capital C has been measured as was intended.

Other goals in gifted programs focus on dependent study or are product- or performance-oriented, and few validated instruments exist to assess such goals. These types of objectives require the development of new measurement strategies which are individualized and often not in paper-and-pencil format.

Measurement problems also exist when evaluators try to assess classroom processes or teacher effectiveness in gifted programs. Training for teachers of the gifted tends to focus on the development of teachers as resources, and on the use of questioning strategies to encourage creativity and thinking skills. Aside from the Aschner-Gallaher System (Aschner, 1959), the Florida Taxonomy of Cognitive Behavior (Brown, Ober, Soar, and Webb, 1967), and the Classroom Activities Questionnaire (Steele, 1981), few instruments exist which assess teacher's strategies with the gifted.

Pitfall 10: Looking for Norms

Control groups, or the difficulty of establishing control groups in gifted programs, present a serious problem to individuals who wish to demonstrate that the "outcomes" are truly attributable to the program and not to other factors. Parents will not be happy to have their children identified as gifted and then left in the control group or regular classroom. One common strategy for dealing with the lack of control groups for students in the programs for the exceptional child is the use of norms as a basis of comparison. That is, one can ask, "What is the 'expected' yearly growth in this population?" and then compare achievements to those norms. Another strategy is to have a control group of identified gifted students who do not get the gifted program treatment immediately, but get the treatment after the dependent variables have been assessed.

Pitfall 11: Avoiding the "Hard" or "Uncomfortable" Issues

A common mistake made by program administrators who evaluate their own programs is to focus only on evaluation questions which they feel will yield positive results, that will not stir up any new issues, and will not reflect on their own performance. To avoid evaluating people, program components, or processes simply because the evaluation process is threatening is detrimental to the overall effectiveness. In fact, when an issue is continually avoided ("Are we identifying the 'right' students?"), that is usually a sign that there *is* a problem and that the evaluator should carefully examine the identification process. Similarly, there is no justification for avoiding the evaluation of any individual's "pet" component. Just because the superintendent suggested the mentor component of a gifted program does not automatically give it protected status. In addition, the credibility of the process is dependent on a sense that *all* components of a program are receiving equal treatment -- including the management and leadership component. If a program is not functioning adequately, it may well be that there is a problem with the management component. Consider this modification of a well-known proverb: "Do not do unto others what you would not have them do unto you."

In planning an evaluation, it is equally important to carry out the process in such a way that individuals do not feel compromised, abused, unfairly treated, or antagonistic toward the program. Tact, fairness, confidentiality, and/or anonymity are crucial in planning, implementing, *and* reporting the results of an evaluation. A situation can be created where decision makers who need to know and act on evaluation information fail to attend to the results because the process has been invalidated in their minds through inappropriate evaluation procedures or evaluator behavior.

Pitfall 12: Expending Resources on Questions to Which No One Really Wants Answers

A corollary of earlier pitfalls, this mistake involves generating questions which will provide potentially useful data and even data that could be considered important data, but data that no one intends to use for decision-making. For example, if staff is being cut in a given school, it is pointless to ask the question, "Is staffing adequate?" in the evaluation.

Pitfall 13: Assuming That Qualitative Evaluation Will Be Easier Than Quantitative Evaluation

When program administrators look at the many problems faced in selecting instruments, setting up appropriate designs, and controlling variables in quantitative evaluation, they are often dismayed and look for alternatives. Often they come to the conclusion that qualitative evaluation designs might best serve their purpose because they are "easier." They are mistaken. Qualitative evaluation designs are, in fact, very rigorous if implemented appropriately; in fact, the appropriate implementation of such designs (connoisseurship, jurisprudential model, investigative model, etc.) is often *more* complex, more time consuming, and more costly than quantitatively oriented designs. In addition, there are few evaluators trained to implement these designs.

Pitfall 14: Assuming That Evaluation Audiences Hear Good and Bad News Equally Well

Human nature leads us to be very anxious about evaluation. We anticipate that any flaw in our performance will be noticed, and are very defensive about negative comments. Accordingly, studies of the communication of evaluation results suggest that negative findings are given more weight than positive information (Locatis, Smith, and Blake, 1980). Baughers (1981) suggests that managers may overreact to negative information and even discount positive information. Neutral comments are often perceived as negatives. He suggests familiarity with the evaluation process seems to result in less overreaction. Thus, it is very important that all of the evaluation procedures be fully explained to administrators and that evaluation results be presented in language least likely to provoke negative reactions and defensiveness.

ADDRESSING THE ISSUES AND PITFALLS

Aside from the obvious ways in which the issues and problems presented thus far may be addressed (clear statement of evaluation purpose, clear identification of audiences, etc.), there are several specific new approaches which the evaluator of gifted programs might consider in improving the effectiveness of evaluations.

The first of these solutions relates to questions of whether the curriculum for the gifted is appropriately differentiated. That is, is the curriculum genuinely different from the regular classroom curriculum? Is it truly beyond the capability of the average ability learner (Carter, 1986)? Does it address topics, themes, and issues of importance and worth for study by gifted students? Does it meet other principles of a differentiated curriculum? Gifted programs are repeatedly criticized for offering options which "everyone could benefit from" or "all students would enjoy" or "all students have a right to benefit from." Carter addressed this issue by expanding a research design suggested by Callahan (1983) to include teaching the same curriculum to gifted students and to a group of average learners as well. An example of his design is graphically illustrated in Table 1. It allows for comparison between the achievement of gifted students studying two different curricular units and comparison of the achievement of average learners as well. An example of his design is graphically illustrated in Table 1. It allows for comparison between the achievement of gifted students studying two different curricular units and comparison of the achievement of average students exposed to the same curricular units. In this design, the evaluator can compare the achievement of the gifted group receiving a higher level thinking skills unit to one receiving the inductive thinking unit with the one group serving as a control to the other. Further, the evaluator can compare the achievement of the gifted group receiving the highest level thinking unit to a regular classroom group receiving the same unit. Presumably, a differentiated curriculum would yield greater achievement for the gifted group. This design allows the evaluator to examine not only the overall effects of curricular units, it also allows for an evaluation of teacher effects.

A new issue which Carter's study highlighted and about which we should be cautious is the "John Henry" effect. When a unit on independent thinking produced significant, but very small, differences between the gifted comparison and treatment groups, further investigation revealed that the comparison group teachers perceived the experiment as a competitive

TABLE 1

Design for Evaluation Gifted Curriculum

CLASS	CURRICULUM UNITS TIME 1*	TIME 2
Gifted Students in Inductive Thinking Unit (IT)		
Gifted Class 1	x	y
Gifted Class 2	x	y
Gifted Class 3	x	y
Gifted Students in Higher Level Thinking Skills Unit (HLT)		
Gifted Class 4	y	x
Gifted Class 5	y	x
Gifted Class 6	y	x
Regular Classroom Students In Inductive Thinking Unit		
Regular Classroom 1	x	-
Regular Classroom 2	x	-
Regular Classroom 3	x	-
Regular Classroom Students In Higher Level Thinking Skills Unit		
Regular Classroom 4	y	-
Regular Classroom 5	y	-
Regular Classroom 6	y	-

*An "x" indicates that a particular group receives the IT unit during that time frame.

*A "y" indicates that a particular group receives the HLT unit during that time frame. Thus, the gifted groups ultimately are instructed using both units while the regular classroom groups receive one or the other unit.

situation. At least one of the teachers permitted twice as much time on the posttest as was allowed the other groups. That teacher also allowed access to materials and resources that were not permitted.

An alternative for improving effectiveness is the adoption of alternative evaluation models such as the case study and/or single subject designs from the exceptional child model of research. Another is to consider creative financing of evaluation to allow for the use of the expertise of professional evaluators. These approaches might include finding ways to incorporate the gifted program as part of larger evaluation studies, cooperative efforts with other school systems, or seeking funds specifically for the development of evaluation designs and instrumentation (See Callahan, Covert, Aylesworth, Crovo, and Vanco, 1980). Seeking ways to share in the development and validation of instruments focusing on product and performance assessment (Reis, 1984; Aylesworth, 1984) would greatly enhance the effectiveness of such evaluations.

And finally, it behooves all to consider the "reasonable person" approach used in the judicial system. That is, does the evidence gathered from many sources, using multiple methods or designs and multiple evaluation instruments or investigators, repeatedly lead us to the same conclusion? If it does, then we can be reasonably assured that those conclusions are tenable and worthy of use in making decisions about program functioning. This process is often referred to as triangulation in the evaluation literature, but too often overlooked in the quest for a research-oriented evaluation design.

Summary

The uneasy political climate in which gifted programs exist mitigates against the widespread implementation of evaluation. The field of gifted education is characterized by cycles of attention and inattention and a constant struggle for special funding to modify programs for a group of students who many still regard as "privileged." Those who have struggled to ensure that "something" is done to meet the needs of the gifted are wary of any process that threatens the existence of these programs or any data that suggest that money is not being used in a cost effective manner. Thus, the summative, judgmental evaluation which has traditionally characterized evaluation of gifted programs has resulted in a clear tendency to avoid the process altogether or to address only the "safe" questions. Gifted programs have failed to carry out evaluations which could have improved programs and helped document the effectiveness of existing strategies to serve the gifted.

One of the fundamental prerequisites of successful evaluation is the same as the prerequisite for creative production -- adequate definition of the problem. In our evaluations of gifted programs, we have often failed to define the problem adequately because we have failed to address the questions of purposes and the questions of audiences adequately. A second fundamental prerequisite of successful creative problem solving is adequate background knowledge and practice in the discipline in which the problem is posed. Our failure to practice evaluation of gifted programs to any significant degree is a serious drawback to finding new and creative approaches to evaluation.

REFERENCES

Aschner, M.J. (1959). "The analysis of classroom discourse; A method and its use." Unpublished doctoral dissertation, University of Illinois, Urbana, IL.

Aylesworth, M. (1984). "Guidelines for selecting instruments in evaluating programs for the gifted." *Journal for the Education of the Gifted,* 1, 38-44.

Baugher, D. (1981). "Summary: Developing a successful measurement program." In D. Baugher (Ed.), *Measuring Effectiveness,* San Francisco: Jossey-Bass.

Brown, B.B., Ober, R., Soar, R., & Webb, J. (1967). *The Florida Taxonomy of Cognitive Behavior: Directions*. Gainesville: University of Florida.

Callahan, C.M. (1983). "Issues in evaluating programs for the gifted." *Gifted Child Quarterly*, 27, 3-8.

Callahan, C.M. & Caldwell, M.S. (1986). "Defensible evaluations on programs for the gifted." In C.J. Maker (Ed.), *Defensible Programs for the Gifted*. Baltimore, MD: Aspen.

Callahan, C.M. Covert, R., Aylesworth, J., Crovo, J., & Vanco, M. (1980). "Evaluating a local gifted program: An LEA/University effort." Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Carter, K.R. (1986). "A cognitive outcomes study to evaluate curriculum for the gifted." *Journal for the Education of the Gifted*, 10. 41-55.

Gallagher, J.J., Weiss, P., Oglesby, K., and Thomas, T. (1983). *The status of gifted/talented education: United States Survey of needs, practices, and policies.* Ventura, CA: Ventura County Superintendent of Schools.

Locatis, C.N., Smith, J.K., & Blake, V.L. (1980). "Effects of evaluation information on decisions." *Educational Review*, 4, 809-823.

Patton, M.Q. (1981). *Creative evaluation.* Beverly Hills, CA: Sage Publications.

Reis, S.M. (1984). "Avoiding the testing trap: Using alternative assessment instruments to evaluate programs for the gifted." *Journal for the Education of the Gifted*, 7, 45-59.

Renzulli, J.S. (1975). *A guidebook for evaluating programs for the gifted and talented.* Ventura, CA: Leadership Training Institute, (Office of the Ventura County Superintendent of Schools).

Steele, J.M. (1981). *Class Activities Questionnaire*. Mansfield Center, CT: Creative Learning Press.

mailing address:

Indiana Department of Education
Office of Gifted and Talented Education
Room 229, State House
Indianapolis, Indiana 46204-2798

office location:

Two Market Square Center